MEASURES
David Rubinow, M.D.

I.   Defining Measures:
   A.   Variable:
         Something to be measured, must be observable and capable of differing or varying (hence variable). Variables can, then, be categorized.
   B.   Scale:
         Definition of the categories (for assessment) of a variable is called scaling. Some variables have well recognized values, while others must have the categories defined by the investigator. The means by which variation among the categories will be assessed is the scale.
      1.   Scale types:
         a)   Nominal - A classification or set of categories that must be "mutually exclusive and collectively exhaustive." Each object always goes into only one category and each can be classified into one (e.g. gender).
         b)   Ordinal - "Mutually exclusive classes that form an ordered series." This is hierarchy or rank ordering. It tells you "moreness" but not how much moreness (e.g. hardness).
         c)   Interval - "The intervals between any two pairs of adjacent classes are equal;" an ordered series of ranks plus equal intervals specifying how much moreness (e.g. temperature).
         d)   Ratio - An interval scale with a true zero point origin.
      2.   Implications of scale type:
         a)   Determines how you summarize the data distribution
         b)   Determines how you analyze differences; i.e. What statistical operations are permissible?
II.   Testing Measures:
   A.   Reliability:
         "The consistency with which a measure assesses a given trait;" i.e., agreement between two measures obtained by the same or "maximally similar" methods (e.g. parallel forms of the same test).
   B.   Validity:
         "The extent to which a measure actually measures a trait;" i.e., agreement between two measures obtained by different methods (e.g. test scores and job performance). What is the relationship between test performance and other independently observable facts about the process under consideration?
         Reliability and validity exist along a continuum, as do "similar" and "different" methods.
      1.   Types of Validity:

a)   Content validity: Are sufficient numbers of the right items included in the right proportions? Is test performance free of the influence of irrelevant variables (e.g. use of difficult or culture-bound words)? (The response as well as the items need to be valid.)   THE MOST IMPORTANT TASK IS TO CONTINUALLY IDENTIFY SOURCES OF RESPONSE CONTAMINATION, VARIANCE OR BIAS.

b)   Criterion-related validity: How well does the measure compare with independent direct measures, either criterion data that are already available (concurrent validity) or subsequent outcomes (predictive validity)?

c)   Construct validity: How well does the test measure the "theoretical construct or trait?" Does performance correlate with measures that it should and not correlate with those with which it should not? If the trait deteriorates with age, does test performance?

Is the test internally consistent? i.e., Do scores on individual test items vary in the same direction as the total test score? Are scores on the test items sensitive (i.e., show change in the predicted direction) to interventions that should change the construct or trait?

Construct validity subsumes all other forms of validity; "there is no information provided by any validation procedure that is not relevant to construct validity."   TEST VALIDITY CAN ONLY BE ESTABLISHED IF YOU KNOW THE LOCAL CONDITIONS (I.E., EVERYTHING THAT MIGHT INFLUENCE OR OBSCURE THE PROCESS YOU ARE ATTEMPTING TO MEASURE) AND IF YOU RESEARCH THE TEST ITSELF.

"It is only through the empirical investigation of the relationships of test scores to other external data" (e.g., variables with which a test score correlates, conditions found to affect performance, and groups in which performance significantly differs) "that we can discover what a test measures."

III.   Measure selection

A.   What do you wish to measure?

1.   Variable domain: disorder-related variables (e.g., symptoms, side effects, correlates of intervention such as drug or hormone levels); disorder-independent variables (e.g. life events).

2. Variable characteristics: symptoms vs. syndrome; severity vs. frequency; absolute value vs. change.
ONE HAS ALWAYS TO CONSIDER (AND MEASURE) THE SOURCES OF NOISE IN YOUR STUDY

B. How do you measure the variable of interest?
1. For measures that already exist: Look at the available data that "tests" the instrument (and ignore its name).
   a) Does it measure what you want and in the proper range given your research question?
   b) What is the sensitivity of the measure?
      (1) "Sensitivity" has two meanings
         (a) Can it discriminate relevant and predictable differences?
         (b) In epidemiology, the false negative rate or proportion of true cases identified

The latter, epidemiological meaning of sensitivity illustrates the importance of knowing how frequently the condition that you wish to detect exists in the population that you are sampling.

The same questions of range and sensitivity (and the construct-related caveats mentioned above) are equally applicable if the investigator creates the measure.

   c) What are the kinetics of the process to be measured? How long and how often must the measure be applied?
   d) What are the factors that may influence the measure?
      (1) Test conditions
      (2) Performance variables
      (3) Lack of sensitivity in the range of interest
      (4) Condition-related and condition-independent factors; in short, the entire range of factors that
   may compromise construct validity.

References

1. Rush, A.J., Gullion, C.M., Raskin, A., Kellner, R., and Bartko, J.J.: Assessment and measurement of clinical change. IN: <u>Clinical Evaluation of Psychotropic Drugs-Principals and Guidelines</u>, R.F. Prien and D.S. Robinson (eds). Raven Press, New York, 1994, pp 85-122.
2. Freeman, L.C.: <u>Elementary Applied Statistics: For Students in Behavioral Science</u>.
3. Schwarz, N., Knauper, B., Hippler, H-J., Noelle-Neuman, E., and Clark, L.: Rating scales numeric values may change the meaning of scale labels. Pub. Opin. Q. 55:570-582, 1991.
4. Schwarz, N.: Judgement in a social context: biases, shortcomings, and the logic of conversation. Adv. Exp. Soc. Psychol. 26:123-162, 1994.
5. Bartko, J.J., and Carpenter, W.T.: On the methods and theory of reliability. J. Nerv. Ment. Dis. 163:307-317, 1976.

6. Bond, A., and Lader, M.: The use of analogue scales in ratings subjective feelings. J. Med. Psychol. 47:211-218, 1974.
7. Anastasi, A.: Chapter 6 Validity: Basic Concepts. IN: <u>Psychological Testing</u>, MacMillan, New York, 1988, pp 134-161.